

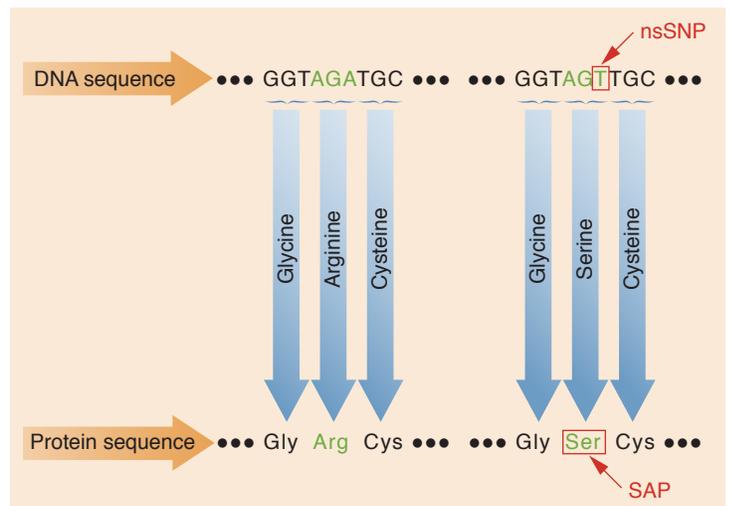
A New Role for Hair in Human Identification



INVESTIGATORS use an assortment of forensic techniques to help identify the perpetrator or victim of a crime, but none are presently considered as reliable as DNA profiling. Only DNA analysis underwent rigorous scientific validation before entering into use by forensic scientists for law enforcement. Moreover, other forensic approaches, such as facial recognition and bite mark, hair, and fingerprint analysis, rely on a specialist to assess how well the evidence matches an image or records in a database, which introduces the risk of human error. A 2009 National Research Council assessment of forensic science detailed the weaknesses of these techniques and reported an urgent need for new science-based methods for human identification that complement DNA analysis.

“Nuclear DNA is the gold standard, but it is quite fragile,” says Brad Hart, director of Lawrence Livermore’s Forensic Science Center. “When the DNA molecule degrades from light, moisture, or heat exposure, it becomes useless for identification.” Even when deterioration is not a factor, some crime scenes—and disaster sites—lack biological material in sufficient quantities or of the right type for DNA analysis. When biological evidence exists but DNA profiling is infeasible, protein analysis could provide a solid forensic alternative.

Proteins are chemically more robust than nuclear DNA and can be found in different tissue types, including hair, shed



This illustration shows a type of genetic code variation called a nonsynonymous single nucleotide polymorphism (nsSNP) and the corresponding altered amino acid in the protein coded by that DNA, called a single amino acid polymorphism (SAP). Forensic scientists can use SAPs (and by extension, nsSNPs) to create an identification profile from protein-bearing tissue samples, such as human hair.

skin cells, bones, and teeth. Livermore forensic scientists and bioinformaticists have teamed up with researchers from Protein-Based Identification Technologies, LLC, to develop the first-ever biological identification method that exploits the information encoded in proteins. Subsequent collaboration with researchers at the University of Utah, Montana State University, University of Bradford, the University of California at Davis, University of Washington, and Utah Valley University supported and provided validation for the approach. The effort aims to identify a person using proteins extracted from a single human hair.

Common Mutations, Uncommon Approach

The new forensic technique looks at genetic mutations through the lens of protein expression. Proteins are long molecular chains formed from amino acids—the 20 basic building blocks of life. DNA is the template the body uses to make proteins. “For a DNA mutation to be reflected in a protein’s amino acid sequence, two things must happen,” explains biotechnology researcher Glendon Parker, founder of the startup company Protein-Based Identification Technologies. “Firstly, the DNA sequence must actually be expressed in a protein. Surprisingly, only one percent of the genetic code actually codes for proteins. Secondly, the mutation cannot be silent, which occurs when two different DNA sequences generate the same amino acid. A tremendous amount of variation exists in humans, so even accounting for these requirements, we still have many candidates to consider.” A DNA mutation that causes a protein to have a different amino acid sequence is called a nonsynonymous single nucleotide polymorphism (nsSNP). The corresponding altered amino acid is a single amino acid polymorphism (SAP).

With support from the Department of Defense, the multi-institutional team has been developing a database of common SAPs—those that appear with one percent or greater frequency in a population—found in human hair samples that can be used to identify a person. Hair currently has limited forensic utility because it lacks nuclear DNA. Within the 400 different proteins the researchers have reliably detected in hair, they have pinpointed 1,700 common locations for mutations and mapped SAPs and corresponding nsSNP variants for 83 of them.

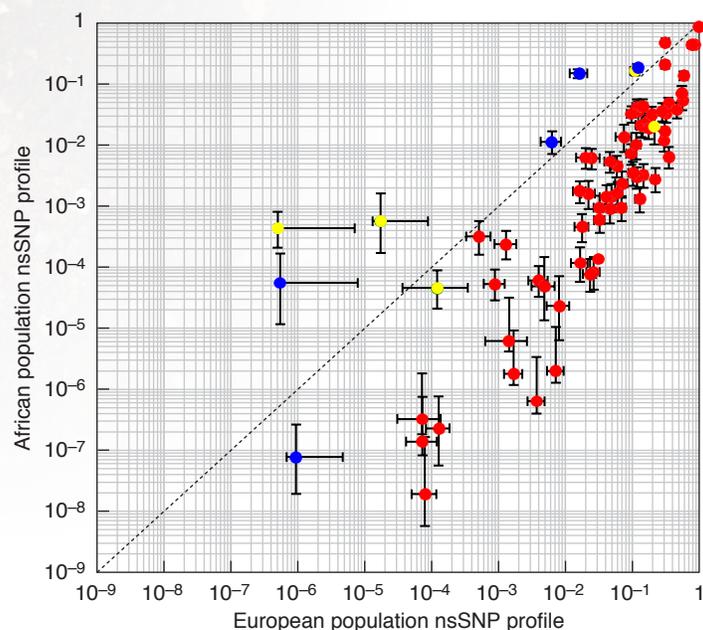
Using this data, the team can perform protein-based identification. The proteins are first extracted from a hair sample and broken down into shorter amino acid chains, called peptides. The researchers then use liquid chromatograph–mass spectrometry (LC-MS) to separate, detect, and quantify the peptide sequences. Results are compared to a sequence database to identify known SAPs present in the sample. Given data on the frequencies of each corresponding nsSNP, the researchers can estimate the power of discrimination for the protein profile. This number increases as more nsSNPs are identified. Livermore’s newly acquired

LC-MS machine has given the team’s work a significant boost. Livermore MS expert Deon Anex notes, “Our state-of-the-art mass spectrometry instrument increases the power of discrimination by two orders of magnitude. We can identify more variants, and we can also obtain the needed information with a tenth the sample size.”

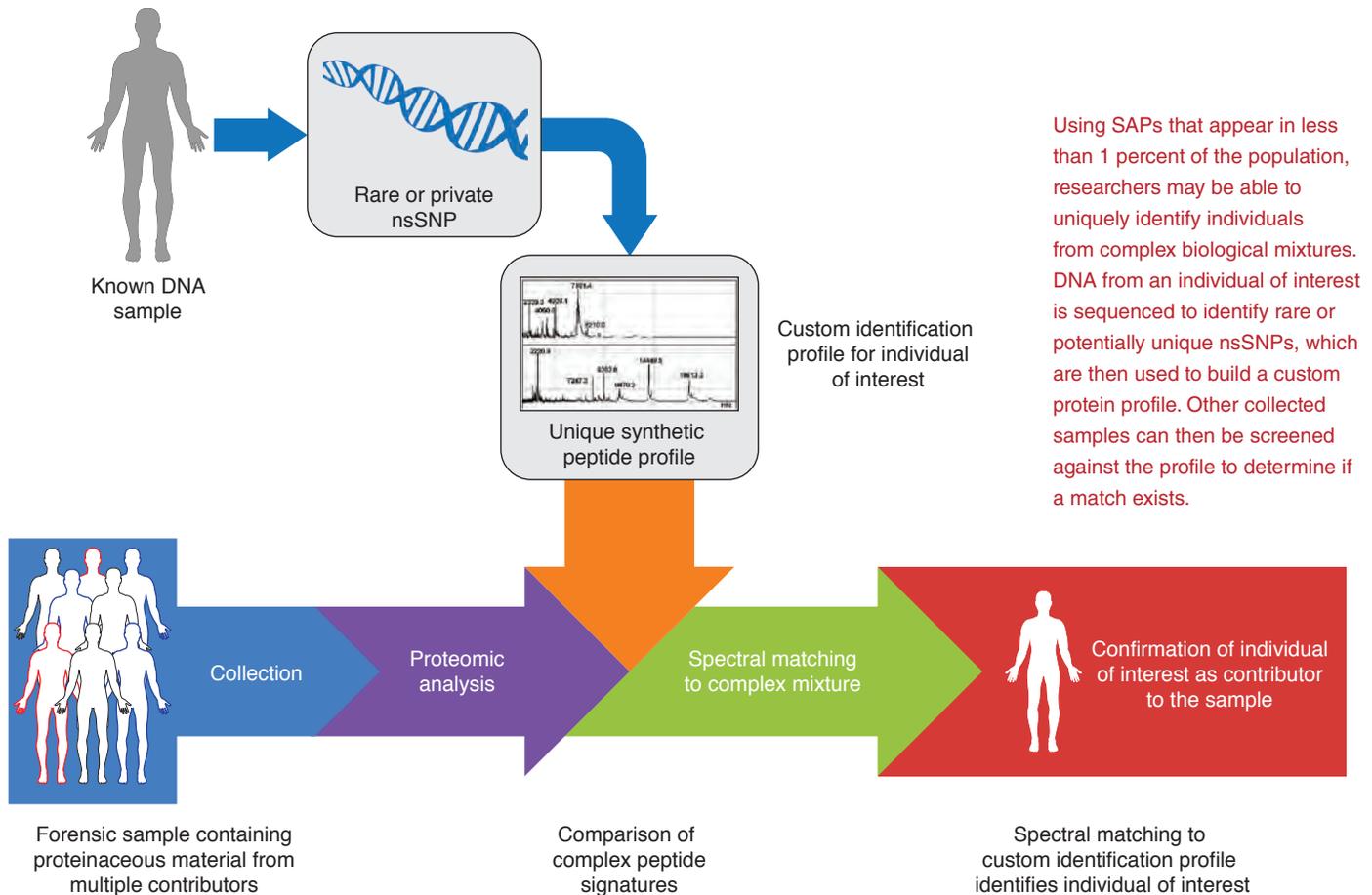
Large-scale events such as human migrations and genetic bottlenecks affect the frequency with which various mutations occur in a population. As a result, the team can also use SAP and nsSNP information to calculate the likely population background of the person who provided a sample. The researchers have tested their approach on European and African genetic pools and plan to expand their database to other genetic groups. This information benefits not only crime solvers but also archeologists. In fact, the researchers have found that they can still reliably discern SAPs and populations of origin from hairs found in 150- to 250-year-old London graves.

Upping the Pace of Research

The team is now investigating less common mutations that will have a higher power of discrimination. Such a capability could be useful for distinguishing individuals within complex biological mixtures, which can be difficult with DNA profiling. Identification



An individual’s likely population of origin can be determined through protein-based identification. The chart shows the probability of various nsSNP profiles occurring in the European and African genetic populations. European–American subjects are marked in red, African–American subjects are yellow, and Kenyan subjects are blue. The dotted line denotes an equal probability of a profile being part of the European and African populations.



using rare SAPs rather than common SAPs requires a more labor-intensive workflow. The forensic scientist first sequences the DNA of the individual of interest or one of the individual's immediate family members and identifies rare or potentially unique nsSNPs in the sample. These nsSNPs are then used to build a custom protein profile against which collected samples can be screened to determine if a match exists. Of course, the accuracy of this method depends on determining the frequency of rare nsSNPs, an area of ongoing research by geneticists.

"This project is occurring at a timely point in genetic research," says Hart. "Currently, 13,000 exomic sequences—the protein-encoding parts of all the genes—are available, mostly of European origin, as are nearly 1,000 genomes of European and African origin. These data sets are rather limited, especially in geographic distribution, but we are at the cusp of a genetic data explosion. By 2018, an estimated one million exomic sequences from a broad geographic distribution will be available to researchers." Access to a vastly expanded library of exomic data will enable Hart, Parker, Anex, and their fellow scientists to quantify more precisely their method's power of discrimination and to pinpoint new mutations.

The team has two major technical hurdles to overcome before protein-based identification enters forensic use. First, the

researchers must further reduce the minimum sample size needed for analysis. Toward this end, they are optimizing protein extraction to ultimately perform identification using just a single human hair. Second, the team has to conduct a statistical analysis on peptides to verify the accuracy of the methods used for calculating the power of discrimination, as these probability estimates presuppose that a SAP in one location is not related to a mutation elsewhere in the protein sequence.

From identifying disaster victims to aiding authorities in catching murderers, protein-based profiling could be a boon to forensic scientists and the broader law-enforcement community. "Twenty-five years ago," says Hart, "DNA identification was in the same place that protein identification is today. This method will be a game changer for forensics, but first we need to prove it."

—Rose Hansen

Key Words: amino acid, exomic sequence, forensic science, Forensic Science Center, genetic research, liquid chromatograph–mass spectrometry (LC-MS), nonsynonymous single nucleotide polymorphism (nsSNP), nuclear DNA, peptide, protein, single amino acid polymorphism (SAP).

For further information contact Brad Hart (925) 423-7374 (hart14@llnl.gov).